LEVEL I

ARI TECHNICAL REPORT
TR-78-B6

AD A077993

ARMY RECRUITERS: CRITERION DEVELOPMENT AND PRELIMINARY VALIDATION OF A SELECTION PROCEDURE

by

George H. Brown, Mark D. Wood, and John D. Harris
Human Resources Research Organization
300 North Washington Street
Alexandria, Virginia 22314

M. A. Fischl, Work Unit Leader
Personnel Accession and Utilization Technical Area, ARI

May 1978

Contract DAHC 19-73-C-0004

Prepared for

U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, Virginia 22333

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

JOSEPH ZEIDNER
Acting Technical Director

W. C. MAUS
COL, GS
Commander

Research accomplished
under contract to the Department of the Army

Human Resources Research Organization

ARI TECHNICAL REPORT
TR-78-B6

ARMY RECRUITERS: CRITERION DEVELOPMENT AND PRELIMINARY VALIDATION OF A SELECTION PROCEDURE

by

George H. Brown, Mark D. Wood, and John D. Harris
Human Resources Research Organization
300 North Washington Street
Alexandria, Virginia 22314

M. A. Fischl, Work Unit Leader
Personnel Accession and Utilization Technical Area, ARI

May 1978

Contract DAHC 19-73-C-0004

Prepared for

U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Army Project Number 20062107A745 Enlisted Accession

This is a report on an exploratory research project designed to meet military management requirements for research on a specific management problem. A limited distribution is made, primarily to the operating agencies directly involved.

ABSTRACT

Research in support of the Army's recruiting operations was conducted to (a) develop a valid criterion of recruiter effectiveness, and (b) develop and evaluate a recruiter selection test battery. Using data from a sample of 400 recruiters, statistical analyses were performed to determine the theoretical yield to be expected from each recruiter's territory based on a multiple correlation between territorial characteristics and production records. A formula was developed to express each recruiter's effectiveness, comparing his actual production with the predicted production. In Task B, tests were assembled to measure recruiter characteristics considered likely to be associated with recruiting effectiveness: verbal fluency, sociability, achievement motivation, empathy, maturity/responsibility, and various background characteristics. The tests were administered to 45 highly successful, and to 43 very unsuccessful, recruiters. None of the individual test scores discriminated significantly between good and poor recruiters. One performance measure of verbal fluency did discriminate significantly, as did about 20 background-information items. The true value of these items for recruiter selection cannot be known until cross-validation has been accomplished.

SUMMARY AND CONCLUSIONS

PROBLEM

With the termination of the draft, the Army's need to maximize the effectiveness of its recruiting operation is clear. The research described in this report was part of a program aimed at developing a procedure for identifying men most likely to be effective recruiters.

OBJECTIVES

Specific objectives of the research were: (a) to develop a valid criterion of recruiter effectiveness, and (b) to develop a selection test (or test battery) for identifying men most likely to succeed as recruiters.

APPROACH

To develop a valid criterion of recruiter effectiveness, a random sample of 400 recruiters was selected. Information was collected on each recruiter's total production (number of accessions) over a six-month period, and on various characteristics of his territory that might influence its fertility. (By "fertility," we mean the relative ease or difficulty of obtaining enlistments in a particular territory.) Using multiple regression techniques, an equation was developed to predict the yield from each territory. Benchmark Achievement Scores (BAS) were then computed to express each man's actual production in relation to the theoretical potential of his territory.

To develop an improved selection procedure, a number of tests—some already existing and some developed in this research—were assembled to measure various characteristics that might be related to recruiter effectiveness: verbal fluency, sociability, achievement motivation, empathy, rejection tolerance, maturity-responsibility, and various background characteristics. Using a composite supervisor rating procedure, 45 of the best recruiters in the Army, and 43 of the poorest, were identified and administered the draft selection test instruments. Results were analyzed to identify items or scores that differentiated between good and poor recruiters.

RESULTS

The criterion development study showed that a single predictor—Average Production per Recruiter in Subject's District Recruiting Command (DRC)—accounted for 48% of the variance in production scores. Average market share (i.e., popularity of the Army compared with the other Services) accounted for an additional 2% of the variance. A Simple Achievement Score (SAS), which expresses each man's production as a percentage of the average for his DRC, correlated +.96 with BAS scores, and was judged to be the preferred measure since it is more easily computed.

In the recruiter selection study, none of the personality measures differentiated between the good and poor recruiters. One performance measure of verbal fluency, the "Ah" ratio, discriminated significantly, as did about 20 background information-type items. These variables have not been cross-validated, however.

CONCLUSIONS

(1) Production scores of recruiters are strongly influenced by the DRC to which they are assigned; in other words, about 50% of the variance in production scores derives from factors unrelated to the individual recruiter's characteristics.

(2) Simple Achievement Scores (SAS) appear to be a more equitable measure of a recruiter's effectiveness than other more traditional measures.

(3) Twenty background items that may be of value in selecting recruiters have been identified, but their true value cannot be assessed without cross-validation.

PREFACE

This report describes research directed at developing an improved procedure for selecting recruiters that will enhance the overall effectiveness of Army recruiting. The research was conducted by the Human Resources Research Organization, under contract with the U.S. Army Research Institute for the Behavioral and Social Sciences, as part of Work Unit RECRUIT, Sub-Unit I, Research on Recruiting.

The work was performed by HumRRO Division No. 7 (Social Sciences) in Alexandria, Virginia. Dr. Robert G. Smith was the Director of this Division, which is now a part of HumRRO's Eastern Division, with Dr. J. Daniel Lyons as Director. Dr. George H. Brown was the Work Unit Leader. Mr. John D. Harris played a major role in developing the draft selection test instruments. SP5 Mark D. Wood, who was assigned by ARI to work full time on this project, performed most of the test administration and assisted in all other phases of the work. The report was written by Dr. Brown and Specialist Wood.

Work Unit RECRUIT was performed for the U.S. Army Recruiting Command. Appreciation is expressed to USAREC personnel for their cooperation, and especially to the individual recruiters who served as research subjects.

HumRRO research for the Department of the Army in Work Unit RECRUIT was conducted under Contract DAHC19-73-C-0004. Army Training Research is conducted under Army Project 2Q062107A745.

Meredith P. Crawford
President
Human Resources Research Organization

TABLE OF CONTENTS

List of Illustrations

List of Tables

Army Recruiters: Criterion Development And Preliminary Validation of a Selection Procedure

Chapter 1

INTRODUCTION

BACKGROUND

With the end of the military draft in January 1973, the Army's need to maximize the effectiveness of its overall recruiting operations was greatly intensified. In June 1972, in anticipation of this need, HumRRO was asked to undertake a research program that would ultimately provide an effective procedure for selecting Army recruiters. The assumption was that good recruiters differ from poor recruiters in some identifiable personal characteristics; if these could be identified and measured, it should be possible to devise a procedure for selecting, as recruiters, only those individuals most likely to be successful at this job.

In the first phase of this project, a pilot study (Work Unit RECRUIT I) was conducted to provide hypotheses concerning the personal characteristics and job behaviors associated with recruiter success. Structured interviews were conducted with a sample of 79 recruiters with high, average, and low records of success in terms of percentage of objective (quota) achieved. Analysis of the pilot study data provided a picture of the nature of recruiting duty and formed a basis for inferring some of the requisite personal characteristics.

REVIEW OF THE LITERATURE ON RECRUITER SELECTION

The research literature on recruiter selection is quite small; much more material is available in the related field of salesman selection. The RECRUIT I pilot study report mentioned earlier contains a short review of virtually all recruiter selection studies published up to that time. That review indicates that no previous effort to develop a recruiter selection instrument had been more than marginally successful.

Among relatively recent studies in this area, three are of interest. Bennett and Haber[1] surveyed a group of 259 Marine recruiters at 29 different recruiting stations to explore the influences on productivity of three classes of variables: (a) characteristics of the individual, such as aptitude test scores and age; (b) geographical characteristics, that is, area of U.S. (e.g., Northeast, Southeast), and (c) deployment variables, such as proximity of the recruiter's working station to his home state. The average number of recruits enlisted per month was used as the criterion of productivity.

Bennett and Haber found that most of their predictor variables had a negligible relationship to recruiter performance and that "the most important determinant of performance is the propensity to enlist in the recruiting market to which the recruiter is assigned. In areas with low enlistment rates, however, recruiters who have served tours as career planners are more productive than others. Moreover, recruiters who work in areas near their home are likely to have an advantage as is a recruiter who works in an urban/suburban environment instead of in a rural area."

[1] J.T. Bennett and S.E. Haber. *Selection, Deployment, and Evaluation of Marine Recruiters*, The George Washington University School of Engineering and Applied Science, Institute of Management Science and Engineering, Project NR 347020, Office of Naval Research, 1973.

As will be seen, Bennett and Haber's emphasis on the importance of territory characteristics finds further support in the present report.

In another study, Abrahams, Neumann, and Rimland[1] used two sample groups of Navy recruiters (Total N=356) to develop a scoring key for the Strong Vocational Interest Blank (SVIB) that appears to be a promising recruiter selection device. Weighted scores were computed for a hold-out group of 178 recruiters who had not participated in development of the scoring key. When these subjects were divided into fourths, the top quarter contained about three times as many "effective" recruiters as the bottom fourth. All subjects had been selected by their supervisors as being among the five most effective, or the five least effective, in their stations. The SVIB appears promising, but, as the authors indicate, further research is needed to determine how well the instrument would discriminate among an otherwise unselected group of recruiter applicants.

In the field of salesmanship research, an interesting study was reported by Grikscheit.[2] Two sample groups of "high effect" and "low effect" salesmen viewed scenes of customer or prospect behavior on closed circuit TV. After each of 16 scenes, the tape was stopped and the subject was asked to describe in writing the verbal and nonverbal behavior he had observed, and to indicate what strategy or tactics he would use next to close the sale. The high effect and low effect salesmen differed principally as follows: (a) The high effect salesmen observed and correctly interpreted more nonverbal cues; (b) they tried a greater variety of strategies early in the interview, but, after settling on one, tended to stay with it. In contrast, low effect salesmen tended to continue changing strategies throughout the interview. Grikscheit indicates that a more comprehensive replication of his study is needed before his findings should be used in selection, training, and evaluation. The extent to which these selling skills are amenable to training, or are identifiable prior to training, is not known. Nevertheless, Grikscheit's work is important and certainly merits serious attention.

THE CRITERION PROBLEM

To attempt to develop an effective recruiter selection instrument is clearly an ambitious undertaking. Because of the following considerations, however, the attempt was made: (a) The need for such an instrument still exists and is even greater since the end of the draft, and (b) most previous attempts, in the fields of both recruiter and salesman selection, have lacked a reliable and valid criterion of effective performance. Any selection study is doomed to failure if the criterion to be predicted is unreliable, or is heavily loaded with variance unrelated to the effective job performance.

Supervisor ratings, which are often used as a criterion in selection research, very often are limited in reliability and validity. Even with the best of intentions, supervisors are often influenced by characteristics that are not truly related to the subject's job effectiveness. For example, a recruiter might be rated high primarily because he is likable, and has a good military bearing and a good production record. Yet the good production record might be the result of having been assigned to a very fertile territory.

School grades (or training course grades) are another criterion often used in selection research. Generally, though, grades reflect the ability to learn verbal materials, rather than to perform on a job. Moreover, they are subject to unintentional grader bias, and often have only a modest correlation with post-training performance on the job.

[1] N.M. Abrahams, I. Neumann, and B. Rimland. *Preliminary Validation of an Interest Inventory for Selection of Navy Recruiters*. Research Memorandum, SRM 73-3, Naval Personnel and Training Research Laboratory, San Diego, California, 1973.

[2] G.M. Grikscheit. "An Experimental Investigation of Persuasive Communication in Selling," paper presented at the American Marketing Association Fall Conference, Washington, 1973.

Raw production figures, while attractive because of their simplicity and obviously high relevance to organizational goals, are a sadly deficient criterion because they are so strongly influenced by opportunity bias. Such figures are a joint function of the individual worker's characteristics and those of the work setting—for example, the fertility of the sales territory, and the quality of the worker's tools and of the management system in which he functions.

The need for a good criterion of recruiter effectiveness is clear. It was made the first order of business in the present study.

OVERALL STUDY PLAN

The overall study plan called for three interlocking steps of data collection and analysis:

(1) Development of a criterion of recruiter effectiveness.
(2) Development and tryout of a set of potential recruiter selection tests.
(3) Revision of the tests and cross-validation on a new sample of recruiters.

Step 1 would be carried out as follows: A random sample of 500 recruiters would be identified and data obtained on their individual production records, their effectiveness as judged by their supervisors, and characteristics which, on *a priori* grounds, might influence the fertility of their territories. It was hoped that the territory data, plus the rating data, would account for most of the variance in production scores. If adequate correlations were found, the territory data would be used to predict the theoretical yield of each territory. Each man would then have a criterion score computed for him that would indicate how well he had performed in relationship to the theoretical potential of the territory. Step 1 was completed on the basis of data on 400 recruiters.

Step 2 called for the assembly of preliminary versions of test instruments likely to measure personal characteristics relating to recruiter effectiveness. The tests were to be tried out on two contrasting groups: 50 of the best recruiters in the country and 50 of the poorest. Items and tests that discriminated between the two groups would be assembled into a revised test battery, which would then be cross-validated in Step 3. Step 2 was completed approximately as planned—the sampled groups consisted of 45 high-rated, and 43 low-rated, recruiters.

Step 3, a cross-validation of the revised test battery, was not accomplished. Only about 20 items were identified that appeared to discriminate between the contrasting groups of recruiters (see Chapter 3). Theoretically these can be assembled into a short paper-and-pencil test that may be worth evaluating in an operational setting, but does not merit a major cross-validation effort. Some suggestions for evaluating the usefulness of this brief test are included in Chapter 3.

Chapter 2

DEVELOPMENT OF A CRITERION OF RECRUITER EFFECTIVENESS

INTRODUCTION

Readily available indices of recruiter effectiveness, such as raw production figures or the percentage of objective achieved, are probably contaminated in varying degrees by "opportunity bias." This term, which comes from the field of industrial psychology, refers to the fact that workers may vary in the amount (or quality) of opportunity they have to produce (make sales). For example, a recruiter assigned to a very fertile (good) territory will probably have a high production record almost regardless of his ability or efforts. The same person, if assigned to a very poor territory, might have a very low production record. The question is: How can a criterion be developed that will not be disturbed by factors outside the recruiter's control?

A preliminary study by Cravens and Woodruff[1] suggested an approach for coping with the problem of opportunity bias. After pointing out that sales volume in a given territory may often indicate the productivity of the territory, rather than of the salesman, they attempted to identify those factors outside the salesman's control that accounted for variance in territory performance (i.e., sales volume). Each factor was defined, measured, and correlated with the criterion of "sales volume." A multiple regression equation was developed and used to predict total sales for each territory. These predicted scores, which were called "benchmark scores," represented the production that a salesman of average ability and motivation might be expected to achieve in that territory.

Next, each salesman's actual performance was expressed as a percentage of his benchmark score, yielding a "Benchmark Achievement Score." Cravens and Woodruff found that the rankings of 25 salesmen in terms of Benchmark Achievement Scores correlated +.61 with their rankings in terms of supervisory ratings. This was statistically significant ($p<.001$), indicating that the benchmark system was reasonably consistent with management's opinions. On the other hand, benchmark rankings correlated only .17 (nonsignificantly) with rankings by quota achievement, suggesting that the benchmark achievement criterion was the more valid in terms of its agreement with management's opinion.

This was the approach to criterion development selected as the model for the research conducted in this study.

ORIGINAL LIST OF POTENTIAL PREDICTORS OF TERRITORY FERTILITY

During an October 1973 meeting, personnel from HumRRO, the U.S. Army Research Institute for the Behavioral and Social Sciences, and the U.S. Army Recruiting

[1] David W. Cravens and Robert B. Woodruff, "An Approach for Determining Criteria of Sales Performance," in *Journal of Applied Psychology*, vol. 57, no. 3, 1973, pp. 242-247.

Command attempted to identify as many factors as possible that might reasonably be expected to influence the fertility of recruiting territories. (Collection of data on some of these factors was not operationally feasible and they had to be omitted from the study.) The factors originally intended for inclusion, and the rationale for including each, are described in the following paragraphs.

Ratio of QMA to MA. Territory boundaries are generally drawn so as to give each recruiter approximately the same Qualified Military Available (QMA), usually 900-1000. Another figure, Military Available (MA), refers to the total male population aged 17-21. in a particular territory. The ratio of QMA to MA indicates the proportion of males in this age group who probably would be acceptable for military service (based on past AFEES rejection rates, etc.). The higher this ratio, the greater the probability that any given applicant will prove acceptable. Thus, this value should be positively correlated with territory fertility.

Geographical Size of Territory. A recruiter with a large territory has a widely dispersed population to deal with. Since he will have to spend a greater amount of time in travel, he will have less time to spend on actual recruiting activities. This factor should be inversely correlated with territory fertility.

Location of Recruiting Station. There are generally four different types of locations for recruiter stations: (a) Post Office or other federal building, (b) suburban shopping center, (c) store front or other congested business area, and (d) professional office building. It is possible that the type of location of the recruiting station influences territory fertility.

Unemployment Rate in Territory. Young men are commonly believed to be more likely to enlist in the Army during periods when civilian jobs are hard to find. If this is true, the unemployment rate should correlate positively with territory fertility.

Median Family Income. It is reasonable to suppose that young men of higher socioeconomic status are more likely to pursue a college education or join a family business than to volunteer for military service. Median family income can serve as an indicator of socioeconomic status and can be indicated by consulting Census Bureau publications.

Educational Level of Community. Presumably, the higher the educational level, the less fertile is the territory as a source of recruits. Information on median school years completed, for each county in each state, is provided in the 1970 U.S. Census.

Racial and Religious Composition of Territory. The racial and religious composition of a territory may possibly influence its fertility. The plan was to seek measures of these characteristics and to examine their relationship with territory production. Possible sources for this information are the U.S. Census reports and the Rand McNally *Commercial Atlas and Marketing Guide*.

Average Market Share. This term refers to the proportion of all military enlistees who, over a specified period of time, chose the Army. It can be regarded as an index of the relative popularity of the Army compared with the other Services.

Average Number of Enlistments per Recruiter in Recruiter's District Recruiting Command (DRC). This number should be a general indicator of the quality of the recruiter's territory, reflecting deep-seated traditions toward military service, as well as complex economic and social factors. This value should correlate positively with territory fertility.

Proximity to a Major Army Post. Recruiter personnel commonly believe that this factor acts to depress enlistment rates. Accordingly, once the sample of recruiters had been selected, a query was to be addressed to each DRC concerned, requesting the name, population of, and distance to the nearest Army installation.

Proximity to a Major Navy or Air Force Installation. This factor is thought by many recruiting personnel to be positively related to Army enlistment rates. Information

on this factor would be solicited from appropriate DRC personnel. Each territory would be categorized as either near or not near a major Navy or Air Force installation.

Proportion of Territory That is Metropolitan, Suburban, or Rural. Statistical analyses of this information—which originally was to be extracted from USAREC Form 100—should reveal whether any one of these territorial compositions is significantly superior to the others.

ASVAB Saturation. The Armed Services Vocational Aptitude Battery (ASVAB) can be administered to civilian high school students at no cost to either the students or their schools. How extensively the battery is used in a given school system is probably influenced both by the receptivity of school officials and the promotional efforts of local recruiting personnel.

Since the degree of ASVAB saturation might reasonably be supposed to be positively related to territory production, information was obtained from USAREC concerning the total number of high school students, within each DRC area, who were administered the ASVAB during FY 74.

Number of 17-21-Year-Olds who are High School Seniors. Presumably the more there are in this age group, the more fertile the territory.

Number of 17-21-Year-Olds who are Enrolled in College. Presumably, this factor would correlate inversely with territory fertility.

Amount of Recruiting Experience. While this is clearly not a territorial characteristic, it is a factor outside the recruiter's immediate control that probably influences the yield of the territory. Since a highly experienced recruiter might reasonably be expected to produce more enlistments than a novice, it is proper to take the recruiter's experience into account in attempting to predict the yield of a territory. (The number of months on production duty as a recruiter served as a measure of experience.)

OBTAINING A SAMPLE OF RECRUITER TERRITORIES

The overall research plan called for selecting a random sample of 500 recruiters' names and compiling territorial information about each. It was originally thought that sample selection and information compilation could be largely accomplished by using USAREC's computerized information system, but some of the necessary information on individual recruiters and on recruiting station territories was not available. USAREC officials then devised an alternate procedure. An official letter was sent from USAREC to the commander of each of the five Recruiting Districts (now called Regional Recruiting Commands) directing him to provide names and other information about 100 individual recruiters. To ensure randomness, names were to be selected by means of the terminal digits of Social Security Account numbers, and were to include only men who (a) had been on production in their current assignment from July to December 1973, and (b) had been assigned individual objectives for that period. (The USAREC letter and the associated data reporting sheet are reproduced in Appendix A.)

Territorial information was compiled by USAREC personnel, partly from available records, and partly by conducting a "special market report."

DESCRIPTION OF THE SAMPLE OF RECRUITER TERRITORIES

Each of the five Regional Recruiting Commands (RRC) was directed by USAREC to supply approximately 100 names, for a total of 500. The number of usable subjects supplied by each RRC is shown in Table 1.

Table 1

Criterion Development Study Sample

Regional Recruiting Command	Number of District Recruiting Commands in Region[a]	Number of District Recruiting Commands in Sample	Number of Recruiters (on Production) in RRC	Number of Recruiters in Sample	Percent of Recruiters in Sample
Northeastern	15	14	905	97	10.7
Southeastern	11	11	772	92	11.9
Southwestern	11	9	687	82	11.9
Midwestern	15	5	1140	66	5.7
Western	10	8	687	63	9.1
Total	66	46	4191	40	

[a]DRCs outside the continental U.S. (San Juan, Honolulu, and Anchorage) were disregarded.

The Midwestern Recruiting Command is clearly not as well represented as the others; only five of its 15 cities are included. However, the present study is a correlational study of the relationship between various territory characteristics and territory production, rather than a descriptive survey. This means that the sample is adequate if each of the predictor variables (territory characteristics) is represented over an adequate range of values and there are no major interactions between place and other characteristics in determining correlations. It seems reasonable to believe that the 10 midwestern cities not included in the sample are essentially similar, on the territorial characteristics of interest, to the 46 throughout the U.S. which are included, and that there are no complex interactions. (A complete list of DRC's, and the number of subjects supplied by each, appears in Appendix B.)

DEFICIENCIES IN THE TERRITORIAL INFORMATION

Ideally, each item of territorial information should refer to the specific geographical area that constitutes a particular recruiter's area of responsibility. However, such fine-grained information was not obtainable within the existing time and cost constraints. Accordingly, most items of territorial information actually apply to the recruiter's station zone rather than to his particular territory. It is likely that in some instances the two are not similar.

The original research design included a verification step that involved contacting by mail the area supervisor for each recruiter in the sample, and asking him to make any adjustments in our territorial information he considered necessary to make it accurate. However, the territorial information became available to HumRRO too late for that step to be taken.

The values for three other potential predictors—unemployment rate, median family income, and educational level of community—were obtained from the U.S. Census Bureau, *County and City Data Book* (1972), in which data are organized only by county or city. This caused problems, since many of the recruiting territories included in our sample comprised portions of several counties. Using county data to calculate census data for such territories would have been time-consuming, and would have involved many dubious assumptions. It was decided to include census data only for those sample

territories whose boundaries were coterminous with county boundaries. Of the 400 cases, 211—concentrated in the densely populated Northeastern Regional Recruiting Command—met this condition.

Most of the data analyses described in the following section were based on the total sample of 400 cases, excluding the three census-type variables. A few analyses were based on the smaller sample of 211 cases on which census-type information was available.

CORRELATION OF PREDICTOR VARIABLES WITH TOTAL PRODUCTION SCORES

The results described in this section were obtained by subjecting 12 predictor variables (excluding the three census-type variables) to a step-wise multiple correlation with "total production scores," that is, the total number of accessions—Nonprior Service (NPS), Prior Service (PS), and WAC's—obtained by each subject from July to December, 1973.

Although in this report we have referred to attempts to predict "territory fertility," our efforts could perhaps be defined more precisely as attempts to determine the extent to which territorial characteristics can be used to predict the *yield* of a territory. This yield can be measured objectively only by the actual production score obtained by the recruiter who worked the territory. Therefore, since our list of potential predictors includes only those variables outside the recruiter's direct control, any predictive accuracy obtained must be attributed primarily to nonrecruiter variables.

A study of the 12 predictor variables used in this analysis, arranged in descending order of their contribution to predicting the criterion of total production, shows that the best single predictor is "Average Production per Recruiter in the Subject's DRC" (Table 2). The correlation coefficient is +.696, which accounts for 48% of the variance in production scores. This means that station (or territory) productivity is strongly influenced by characteristics of the particular station, which, in turn, could be social or cultural features not measured in the present study. (None of the social or cultural features measured in this study were found to relate strongly to productivity.)

It is also possible that "Average Production per Recruiter in the DRC" is itself a function of local personnel management practices, and of the quality of the individual recruiters. Whether any selective factors operate in the assignment of recruiters to DRC's is not known.

It should be noted that the large value of this correlation (+.696) is to a small extent artifactual, since the recruiters whose production records constituted the criterion were themselves contributors to the average production of their DRC. However, since our total N of 400 is only about a 10% or smaller sample of all recruiters in the United States, this is not a serious contamination.

It remains true, as indicated by the correlation coefficient (+.696), that recruiters tend to produce at a rate that is characteristic of their DRC.

The second best predictor is "Average Market Share for Station Zone" (Table 2). This variable can be viewed as an index of the relative popularity of the Army, as compared to the other services. When this variable is added to the prediction equation, the multiple R increases to +.710, and the amount of criterion variance accounted for rises to 50.5%. A third variable, "Proportion of Zone Which is Suburban," adds a minute amount to the multiple R, which then becomes .714.

The remaining variables listed in Table 2 add only trivially to predictive power. The reasons why several potential predictors which, on *a priori* grounds appeared promising, do not rank high can be seen by studying the "Simple r" and "Partial r" columns. The values in the "Simple r" column are the simple correlations between the predictor

Table 2

Stepwise Multiple Correlation (R) of 12 Predictor Variables With Total Production Over a Six-Month Period

(N = 400)

Predictor Variables	Simple r[a]	Partial r[a]	Successive Multiple R[a]	Proportion of Variance Accounted For
Average Production per Recruiter in Subject's DRC	.696*	.389*	.696*	.485
Average Market Share for Station Zone	.549*	.202*	.710*	.505
Proportion of Zone Which is Suburban	-.092	.058	.714*	.509
Months of Experience as Recruiter	-.012	.09	.716*	.512
Number of High School Seniors in Zone	-.054	.041	.717*	.514
Average Production per Recruiter for Subject's RCC	.595*	.027	.717*	.515
Number of ASVAB's in Subject's DRC	.173*	.029	.718*	.515
Number of 17-21-Year-Olds in College in Station Zone	-.040	.018	.718*	.515
Size of Station Zone in Square Miles	.096	-.011	.718*	.515
Proportion of Zone Which is Rural	.126*	.069	.718*	.515
Proportion of Zone Which is Metro	-.034	.075	.720*	.518
Ratio of CMA to MA	-.040	.007	.720*	.518

[a]*indicates statistical significance ($p < .05$).

variables and the criterion. While several of these are significantly larger than zero, they do not appear important in the multiple R because they were largely duplicating other predictors with which they were correlated. For example, "Average Production per Recruiter for Subject's RRC" correlated +.59 with the criterion of total production; but since it also correlated highly ($r = .820$) with "Average Production per Recruiter in Subject's DRC," it made essentially no unique contribution to prediction and ranks only sixth among the predictors.

The column of partial r's shows the correlation of each predictor with the criterion when all other variables are held constant (neutralized) by statistical means. The fact that most of the partial r's are small and nonsignificant is consistent with the fact that including them in the multiple regression would add little to predictive power.

Now let us consider the results obtained when all 15 potential predictors are put into the regression equation, that is, when we add the three census variables, and necessarily reduce the N to the number for whom census data are available (N = 211). In a word, prediction worsens; the R shrinks to .59 and accounts for only 44% of criterion variance. This shrinkage probably results from the fact that the subset of recruiter territories is more homogeneous than the population as a whole, thereby producing a "restriction in range." As indicated earlier, the Northeastern Regional Recruiting Command contributed a disproportionate share of subjects for this analysis (about 40% of the 211 cases).

CORRELATION OF PREDICTOR VARIABLES WITH OTHER CRITERIA

The kinds of analyses described in the previous section were replicated using two other criteria: (a) percent of objective achieved over a 6-month period, and (b) total point score[1] over a 6-month period. The maximum R obtained using 12 variables and all 400 cases was +.619 for the former, and +.733 for the latter. Evidently, total point score is more predictable than total production score.

In all three analyses, the highest ranking predictors were the two previously described: (a) "Average Production per Recruiter in Subject's DRC," and (b) "Average Market Share for Station Zone."

COMPUTATION OF BENCHMARK ACHIEVEMENT SCORES (BAS SCORES)

BAS scores were computed by the following formula:

$$\text{BAS Score} = \frac{\text{Actual Production}}{\text{Predicted Production}} \times 100$$

"Actual Production" scores, which were supplied by USAREC, represent the total number of accessions to the Army produced by each subject from July to December, 1973.

Predicted production scores were computed via the multiple regression equation represented in Table 2. Only the top three predictor variables were used, since together they account for approximately 51% of the variance in production scores, and adding the other nine variables would increase the amount of variance explained by less than one percentage point. The three variables were:

(1) Average Production per Recruiter in Subject's DRC.
(2) Average Market Share for Station Zone.
(3) Proportion of Zone Which is Surburban.

The distribution of BAS scores is presented in Figure 1. The mean score is 99.1 and the range is from 7.1 to 245.9. The general appearance of the distribution suggests normality—that is, the curve is basically bell-shaped. In a perfectly normal distribution, 68% of the cases would be within one standard deviation of the mean, that is, between 64.0 and 134.2. Results in this distribution are close to that, with 73% of the cases falling within these limits.

VALUE OF BAS SCORES AS A CRITERION OF RECRUITER EFFECTIVENESS

The value of BAS scores must be assessed primarily on rational or judgmental grounds. The original research design for this criterion study called for collecting sophisticated ratings of the overall effectiveness of each recruiter in the sample from five different supervisors, but there was not enough time to take this step. If such ratings were available, it would be interesting to find out how much of the variance in territory production they would account for—and to determine how they correlate with BAS scores. A higher correlation than that between ratings and production scores would

[1] As an incentive, recruiters are awarded varying numbers of "points" for each accession—the number of points varying with the "quality" of the accession. For example, more points are awarded for enlisting a high school graduate than a non-high school graduate.

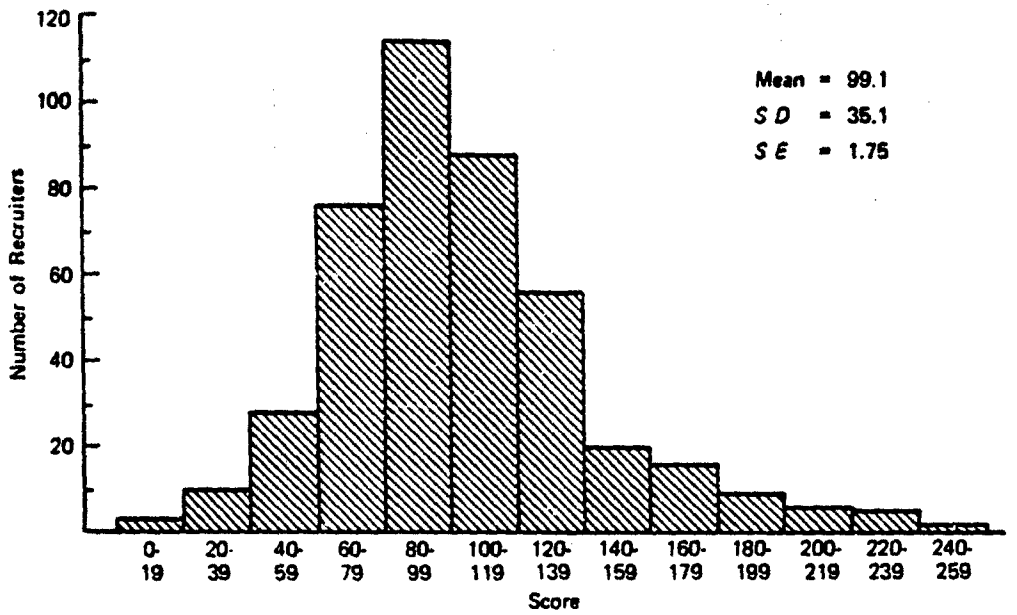

Figure 1 **Distribution of Benchmark Achievement Scores of 400 Recruiters**

suggest that BAS scores are a more valid indicator (criterion) of recruiter effectiveness than either production scores, or, perhaps, scores based on percentage of objectives achieved.

To illuminate the meaning and possible validity of BAS scores, their correlations with other available indices of recruiter effectiveness were computed (Table 3). Since BAS scores were designed specifically to evaluate a recruiter's performance in relationship to the quality of his territory, it is not surprising that they correlate only .646 with raw production scores. They also correlate rather low with "percent of objective scores" and with total point scores. Clearly, BAS scores do not duplicate the functions of these other scores, which suggests that they are indeed giving due weight to variations in quality of territory.

Table 3

Intercorrelations Among Various Indices of Recruiter Effectiveness, July - December 1973

(N = 400)

Indices of Recruiter Effectiveness	Total Production	Percent of Objective Score	Total Point Score	SAS[a] Score
BAS Score	.646	.619	.564	.957
Total Production		.847	.929	.729
Percent of Objective			.818	.701
Total Point Score				.647

[a]Simple Achievement Score.

SIMPLE ACHIEVEMENT SCORES (SAS)

The SAS scores that appear in Table 3 have not been mentioned previously. It occurred to the authors that, since the predictor variable of predominant importance in computing BAS scores is "Average Production per Recruiter in Subject's DRC," a Simple Achievement Score (SAS) based solely on the individual's performance compared to that average might be equally suitable—and much easier to compute. Accordingly, for each subject in the sample, an SAS score was computed as follows:

$$\text{SAS Score} = \frac{\text{Total Production}}{\text{Average Production in DRC}} \times 100$$

This formula produces SAS scores that express each man's production as a percentage of the average for his DRC. As Table 3 shows, SAS scores correlate extremely high with BAS scores ($r = .957$). Thus, the two scores are essentially equivalent.

As a criterion of recruiter effectiveness, the SAS score appears to be somewhat preferable to the BAS score simply because it is more easily computed. But is it a good criterion measure? This question can only be answered judgmentally. The SAS score clearly provides a wide range of variation from one recruiter to another, that is, it has good discriminating power. It is intuitively meaningful, since, in effect, it rates each man against the average for his DRC, thus taking into account a host of unknown factors that make some territories better than others. One would surmise that SAS scores would be generally acceptable to the recruiters themselves.

It is hoped that future research can be done in which our recommended new criterion measures (SAS scores) can be correlated with selection test scores. If the SAS score can be predicted more accurately than older criteria, this fact will itself support its validity or relevance.

Chapter 3

DEVELOPMENT AND TRYOUT OF A RECRUITER SELECTION TEST BATTERY

GENERAL PLAN

In developing a recruiter selection test battery, personal characteristics that might reasonably be expected to contribute to effectiveness as a recruiter were identified on the basis of: (a) examination of, and deductions from, the pilot study findings; (b) conversations with personnel from Headquarters, USAREC, and the Washington Recruiting Main Station; and (c) common sense considerations. In brief, the characteristics to be measured were: verbal fluency, sociability, achievement motivation, empathy, rejection tolerance, maturity/responsibility, and a variety of background and demographic features.

The general plan for evaluating the tests called for administering them to two groups composed, respectively, of 50 highly successful, and 50 very unsuccessful, recruiters. Items or tests that discriminated between these contrasting groups would be assembled into a revised (and shorter) selection test, which would then be cross-validated on a new sample of recruiters. The criterion to be used in the cross-validation study would be that developed according to the procedures described in Chapter 2.

DESCRIPTION OF THE TEST INSTRUMENTS

In physical format, the test battery consisted of 12 test booklets and one verbal performance test. Some tests measured more than one characteristic and some characteristics were assessed by more than one test. A variety of formats were used for the test items, including true-false, multiple choice, extent of agreement or disagreement, and open-end write-in items. In addition to the information obtained from the tests, a few items of information were obtained from each subject's official personnel form (e.g., aptitude test scores, race, religion).

The descriptions of the test instruments that follow are organized around the characteristics the tests were intended to measure.

Verbal Fluency Measures

It seems obvious that an effective recruiter must be able to talk easily in a variety of social situations. "Effective Speaking" is one of the characteristics on which applicants for recruiting duty are evaluated (subjectively) by recruiter selection boards. It would seem that an effective recruiter must be able to express his thoughts readily, if not necessarily grammatically.

Most existing tests of verbal fluency are written tests requiring the subject, within a short time, to name as many words as he can that start with a given letter or pertain to a given semantic area. However, the authors believe that verbal fluency should be measured orally, in the most realistic situation possible. Accordingly, each subject was asked to speak into a dictaphone "mike" and pretend that he was making a sales pitch to a potential enlistee who knew little about the Army.

The test administrator ad libbed the instructions to the subjects, somewhat as follows:

> You are being asked to speak for three to five minutes on general Army benefits as if you were speaking to a prospect. We realize that your presentation to a prospect in a real situation varies, depending on your initial assessment of the individual. However, our purpose is to find out how recruiters differ in their presentations, so we simply ask you to speak on general Army benefits. Imagine that I (the administrator) am a prospect interested in the Army but unfamiliar with the benefits and or options available to me should I enlist.
>
> When you are ready, say so, and we will begin.

After each subject had spoken for at least two minutes and reached a logical stopping point, the recording was stopped. Occasionally a subject would stop short of two minutes, in which case the administrator would simply terminate the exercise.

A few subjects obviously had great difficulty in carrying out the assigned task; they complained that the situation was unnatural. In such instances, the administrator repeated the instructions, elaborating on them with more detail, and the subject then complied. The vast majority of the subjects grasped the idea quickly and performed without difficulty.

The first two minutes of each recording were transcribed into type for scoring purposes. The original plan was to develop a variety of scores for each transcript, following a procedure developed by Preston and Gardner.[1] However, because of time constraints, only two rather simple scores could be computed. These were:

- Verbal A: Total number of words produced. For those few subjects who stopped short of two minutes, their scores were extrapolated to obtain a value corresponding to what they would have presumably scored in two minutes.
- Verbal B: "Ah" Ratio. The total number of "ah's" uttered was divided by the total number of words uttered. Presumably, a more verbally fluent person would have a lower Ah ratio.

Sociability Measures

Sociability, or the tendency to enjoy interacting with people, seemed important since a recruiter must spend much of his time interacting with people, who often are strangers. If he did not receive a certain satisfaction from these encounters, he would probably be underactive in prospecting and deficient in sales.

Four of the 13 tests in the battery were intended to measure sociability. Each of these is described.

Personal Preference Test No. 1. This test consists of 11 items, each of which describes a hypothetical activity that could be performed either alone or with various kinds of companions (wife, friend, acquaintance, several people). The subject reads each activity and assigns rank numbers, 1 through 5, indicating his order of preference for the different ways of performing the activity. A sample question reads:

> If I were going to a party, I would prefer to go:
> ____ a. with my wife girl friend
> ____ b. with an acquaintance
> ____ c. alone
> ____ d. with a friend
> ____ e. with several other people

[1] Joan M. Preston, and R.C. Garner, "Dimensions of Oral and Written Language Fluency," in *Journal of Verbal Learning and Verbal Behavior*, vol. 6, December 1967, pp. 936-945.

Four of the five preferences are of a social nature, the preference "alone" being considered asocial. A subject who repeatedly assigns "5" to the "alone" preference shows a high level of sociability; if he repeatedly assigns "1" to "alone," a low level of sociability is indicated. The test is scored by summing the ranks assigned to "alone" and subtracting 14. Thus, scores can range from zero to 56.

Personal Preference Test No. 2. Similar to the previous test in intent, this test consists of 22 items, only 11 of which are scored. All items require the subject to express his preference about companionship when performing a certain activity. For the 11 scored items, the choice is between performing the activity alone versus doing it with others. Scores could range from zero to 11. A sample item reads:

a. Go on a weekend trip with some other people
b. Go on a weekend trip alone.

The subject is asked to take into account what the activity is and with whom he would be doing it. Although individual activities are repeated, no two pairs of activities are the same.

Personality Inventory. The title "Personality Inventory" is an arbitrary label for a test consisting of slightly modified versions of two tests developed by Dr. Albert Mehrabian, of the University of California, Los Angeles.[1] Thirty-four of the items in the Personality Inventory are actually Mehrabian's items for measuring affiliative tendency, which we refer to as "sociability." The subject must indicate his degree of agreement or disagreement on a 5-point scale (+2 [strongly agree], +1, 0, -1, -2 [strongly disagree]) to statements such as: "When I'm introduced to someone new, I don't make much effort to be liked," and "When I'm not feeling well, I would rather be with others than alone."

To obtain a perfect score, half of the 34 items must be marked +, to indicate sociability; the other half must be marked minus. Items marked either +2 or -2 indicate maximum sociability. For a total score, the minus sign for each of the subject's negative responses for a negative item must be changed to a plus and all responses summed. Scores can range from zero to 68.

Social Activities Questionnaire. This questionnaire attempts to determine the extent of the subject's social life and the kinds of activities he enjoys. Nominally, there are 17 items, but several of them have scorable components. Topics dealt with include: type of residence occupied, length of residence in the present neighborhood, number of neighbors considered close friends, frequency of visits to friends, and frequency of social functions such as parties, movies, night clubs. In addition to queries about how often he engaged in these activities, the subject was asked how often he would *like* to engage in them.

Achievement Motivation Measures

Achievement motivation, or the urge to work hard to achieve self-appointed goals, is thought to be a positive characteristic of a good recruiter. Such a person would be thought of as hardworking, industrious, persevering, and energetic. The following three tests were designed, either wholly, or in part, to measure achievement motivation.

Personal Opinion Survey. This test consists of 35 maxims purporting to express a philosophy related to achievement motivation. They are simplified statements made by various famous Americans, all of which say something about life or men in general. Using a five-point scale (+2, +1, 0, -1, -2) the subject indicates his degree of agreement or

[1] Permission to use these tests on an experimental basis was granted by Dr. Mehrabian. In their original form they were called measures of Affiliative Tendency and of Male Achieving Tendency. For a complete description of the original tests, see A. Mehrabian. "Measures of Achieving Tendency," in *Educational and Psychological Measurement*, vol. 29, 1969, pp. 445-451 and A. Mehrabian. "The Development and Validation of Measures of Affiliation Tendency and Sensitivity to Rejection," in *Educational and Psychological Measurement*, vol. 30, 1970, pp. 417-428.

disagreement with each maxim. A sample item reads: "What a man does, that he has. In himself is his might."

Personality Inventory (even numbered items, 2-52). These 26 items constitute a slightly modified version of Mehrabian's measure of Male Achieving Tendency. For a complete description of the original test, see Mehrabian.[1,2]

Background Information Forms. Based upon ideas expressed by McClelland,[3] a number of biographical items were devised which, according to McClelland's findings or theories, might be expected to measure achievement motivation. The items inquired about such matters as the nature and extent of the subject's participation in competitive sports, the kinds of jobs he had held, his attitudes toward a variety of high status jobs, and the educational level of his parents.

The test itself consisted of 40 multiple-choice and multiple-check items which were scored empirically. That is to say, there were no predetermined right or wrong answers. The intent was to determine empirically whether any items were answered differently by good and poor recruiters.

Empathy Measures

Empathy is defined here as the ability to understand the point of view of others and to correctly perceive the impact one is making on another. McMurry[4] has argued convincingly that the super-salesman possesses a high degree of empathy, coupled with a high degree of what he calls "ego-drive" (the urge to win). In other words, the super-salesman must be capable of fully grasping the prospect's point of view but, to be able to close the sale, he must not permit excessive sympathy to neutralize his urge to win.

Four instruments were used in the attempt to measure empathy.

Knowledge of Prospects Test No. 1. Both of the Knowledge of Prospects tests are based on the rationale that a recruiter who is accurately informed about the political and social attitudes of today's young men will be more empathic with them. He should understand them better and be more effective in influencing them to enlist in the Army.

Information obtained from the Purdue Poll No. 89 (1970), which surveyed American high school students (grades 10-12), resulted in 22 True-False items and 40 multiple-choice items concerned with student thoughts on prejudice, poverty, peace, population, pollution, and so forth.

Part I of this instrument asks the recruiter to respond—True or False—to 22 statements describing "the typical male high school senior." If he thinks the statement is more often true than not, he is to mark True. Otherwise, he marks False. A sample item is:

> "He is primarily interested in going to college after he finishes high school and has little if any interest in entering military service."

Part II of the instrument gives two brief, general descriptions of hypothetical individuals: Fred Thompson, a 17-year-old, white, middle-class high school student living in suburbia; and Pete Kerber, a black, unmarried, 19-year-old, inner-city man who has graduated from high school and is now working as a drug store clerk in the suburbs. The

[1] A. Mehrabian. "Male and Female Scales of the Tendency to Achieve," in *Educational and Psychological Measurement*, vol. 28, 1968, pp. 493-502.

[2] Mehrabian, 1969, *op. cit.*

[3] McClelland, D.C. *The Achieving Society*, Van Nostrand, Princeton, N.J., 1961.

[4] R.N. McMurry. "The Mystique of Super-Salesmanship," in *Harvard Business Review*, vol. 39, no. 2, March-April, 1961.

recruiter's task is to answer identical sets of multiple-choice questions concerning these two young men. A sample item is:

> "Which of the following incentives would probably be most effective in inducing_____to enlist in a combat branch?"
> a. a large cash bonus
> b. a paid college education
> c. military pay comparable to civilian pay
> d. guaranteed assignment in a certain military specialty

The test is scored in terms of the number of items marked correctly. Possible scores can range from 0 to 57.

Knowledge of Prospects Test No. 2. After Test No. 1 was assembled, a more recent poll of a similar nature came to the authors' attention: "A Survey of Attitudes and Motivations Towards Enlistment in the Volunteer Army" (Opinion Research Corporation, 1974). Hence, Knowledge of Prospects Test No. 2 was created, based on the more recent, and presumably more valid poll.

The more recent survey collected the opinions of (a) American males, aged 14-21; (b) the parents of young men aged 17-21; and (c) educators from various high schools. From the published results, questions were formulated regarding the non-college males, aged 17-21, who constitute the prime enlistment prospects. The following is a sample of the 30 True-False items created:

> "He considers time spent in the Armed Forces as time ill-spent, as time wasted."

Personnel Questionnaire. The title "Personnel Questionnaire" is simply an arbitrary label for a collection of background-type items, 11 of which might measure empathy. Sample content areas were: number of siblings, and how often friends seek the subject's advice. Scoring was empirical, by individual item.

Hogan Psychological Inventory. This test was actually the Hogan Empathy Scale.[1] It consists of 64 True-False items, of which the following are examples:

> T F I would certainly enjoy beating a crook at his own game.
> T F I frequently undertake more than I can accomplish.
> T F I prefer a shower to a bath.

Rejection Tolerance Measure

A recruiter who prospects actively will inevitably receive rejections, rebuffs, and occasional insults. If he typically reacts to such experiences with distress and discouragement, both his prospecting effectiveness and his production will probably deteriorate. It seems likely that highly successful recruiters would have a higher tolerance for rejection than the less successful. The Social Interactions Test was designed to measure this characteristic.

The Social Interactions Test. Based on a factor description developed by Mehrabian[2] a 53-item True-False test to measure rejection tolerance was created. Mehrabian's studies of affiliative tendency and sensitivity to rejection supply the notion that persons characterized as being high on these two scales exhibit less flexible

[1] Used with the permission of Dr. R. Hogan. For a detailed description, see R. Hogan. "Development of an Empathy Scale," in *Journal of Consulting and Clinical Psychology*, vol. 33, no. 3, 1969, pp. 307-316.

[2] Mehrabian, 1970, *op. cit.*

interpersonal behaviors. They are much more likely to conform and are less willing to associate with people whose attitudes and opinions differ from their own. Sample items from the test are:

T F I try to avoid talking to people who seem cool and stand-offish.

T F When I know a person is the type that makes critical comments, I avoid talking to him.

The scoring system developed for the test was consistent with Mehrabian's factor description. Scores could theoretically range from 0 to 53.

Responsibility-Maturity Measures

Since a recruiter spends the bulk of his duty time working without supervision, and since he represents the Army to the general public, it is essential that he manage his personal, financial, and official duties with discretion. Responsibility is a characteristic to which recruiter selection boards pay particular attention. The three instruments described in this section were used to measure this characteristic.

Background Information Form (Items 22, 23).

Personnel Questionnaire (Items 14-19).

Various items in these two instruments attempt to measure responsibility or maturity. Sample items deal with the purposes for which the subject has borrowed money, and the number of automobile accidents in which he has been involved. Scoring was empirical.

The I-E Scale. Rotter's I-E Scale (Internal External)[1] is a questionnaire designed to assess the degree to which a person thinks reward (reinforcement) is contingent on his own behavior (internal control) or independent of it (external control). Presumably individuals with a strong sense of responsibility would score high on the Internal Control dimension.

The questionnaire consists of 29 items, each of which provides alternative statements about the ways people react to events in society and in their personal lives. The subject selects the statement in each pair that he believes more strongly to be the case. He is specifically instructed to select the statement he actually believes to be true, rather than the one he thinks he should choose or would like to be true. A sample item is:

a. Many of the unhappy things in people's lives are partly due to bad luck.

b. People's misfortunes result from the mistakes they make.

Scoring was in accordance with Rotter's key. Six of the 29 items are fillers (i.e., are not scored). Scores can range from 0 23. Strictly speaking, scores represent the strength of the subject's belief in external control; low scores thus imply the opposite.

SELECTION OF RESEARCH SUBJECTS

Since the criterion development effort was not completed in time to use the new criterion scores as a basis for evaluating the potential selection tests, the decision was

[1] Reproduced by permission of Dr Julian Rotter. See J.B. Rotter, "Generalized Expectancies for Internal Versus External Control of Reinforcement," in *Psychological Monographs*, no 609, vol 80, no 1, 1966

made to evaluate the tests by administering them to two groups of recruiters, representing the extremes of talent. The plan called for selecting 50 of the best recruiters in the country and 50 of the poorest, and determining whether the test instruments would discriminate between them.

To ensure adequate geographical representation, two DRCs were selected from each of the five RRCs (formerly called Districts). Those selected were the highest and the lowest ranking DRCs within each RRC, based on the percentage of objectives achieved from July to December, 1973.

The following procedure was used to identify for testing the five highest and the five poorest recruiters within each DRC. Each of five supervisory personnel was asked to nominate in writing the *10 best* and the *10 poorest* recruiters within his DRC. (The nomination form is reproduced in Appendix C.) For each name that appeared in either list, a tally was made of the total number of times he was mentioned. The five men mentioned most often on the "best recruiter" list were selected as the subjects in our High Criterion Group. The five mentioned most often on the "poor recruiter" list were designated as our Low Criterion Group.

Although this plan involves the use of supervisor ratings, which we have characterized as generally unreliable, it nevertheless appears to be sound. Using the composite judgment of five supervisors, instead of one, should provide more reliable ratings. It seems safe to assume that the two extremes of recruiter talent can be identified in this way. Also, the comparison of two such extreme groups should maximize the chances of identifying the characteristics that relate significantly to recruiter effectiveness.

However, a few problems arose in carrying out the plan. At two DRCs, the "poor recruiter" lists were not filled out completely. When queried, the supervisors stated that even their poorest man was consistently exceeding his objective—which makes it clear that a poor man in one DRC might well surpass a good man in another DRC. Since time did not permit additional efforts to complete the lists, subjects were selected from the lists available.

Also, when the time came for scheduling the testing sessions, many of the designated subjects were not available for one reason or another, such as leave, illness, or transfer. In such cases, alternates were selected from the original lists of nominees. The number of Highs and Lows actually tested was 45 and 43 respectively.

ADMINISTRATION OF THE TEST BATTERY

The test battery was administered from 22 April to 21 May 1974 by Mark D. Wood. Approximately one month before testing began, each participating DRC received a communication from USAREC describing the general purpose of the project and the support required. Subsequently, the test administrator telephoned his DRC contact to make detailed arrangements regarding date, time, physical facilities, and subjects. Later, at the DRC, he briefed the commander (or his representative) on the nature of the projects and arranged to obtain production records and other information on the test subjects.

The physical facilities for testing were generally satisfactory: a quiet room with convenient writing surfaces and another quiet place for recording the verbal performance test. There were some problems, however. For example, of the 10 DRCs visited, three could not bring all 10 recruiters together for a single administration, so the test battery had to be administered at two different locations. In one of these sites a busy recruiting station testing was subject to frequent interruptions and distractions. At another DRC, where the commander stipulated staggered test periods

to avoid taking too many men off production at once, a group testing situation was not achieved. The effect on test performance cannot be ascertained.

As part of their introduction to the testing, the subjects were informed that USAREC was interested in developing an improved selection/screening procedure for identifying potential recruiters, in order to avoid the waste of time and effort involved in training men who would not succeed in the job. They were also told how long the test would take (4-6 hours was the initially stated time, but it seldom took more than 4½ hours, including a break for lunch).

The men were not told precisely how they had been selected. Whenever the question arose, it was explained that their names had been supplied by DRC administrative personnel. In general terms, the men were told that we were trying out the tests—seeking to find out whether people with different production levels would answer the test questions differently. Without exception, the cooperation of the test subjects was excellent.

DATA ANALYSIS

Data analysis was directed at discovering significant differences between the High and the Low criterion groups. Most of the tests with existing scoring keys were hand-scored; each subject's score was recorded on a "Record Sheet," along with certain data obtained from Army personnel records, such as, race, religion, GT and CL scores. Record sheet data were then recorded on punch cards, along with individual item responses to the other tests.

Analyses were relatively straightforward. For all test scores, and for all item responses that were continuous variables, means and *t*-tests were computed. For all other items, frequencies and percentages of the High and Low criterion groups who marked each response option were computed. Whenever the percentages of the two groups appeared to differ substantially, a two-by-two chi-square statistic was computed to evaluate the statistical significance of the difference.

Chapter 4

EVALUATION OF TEST BATTERY

RESULTS AND DISCUSSION

None of the personality measures, and none of the scores extracted from Army personnel records, discriminated significantly between the High and Low criterion groups. The actual mean scores are presented in Table 4.

Table 4

Mean Scores of High and Low Criterion Groups on Various Personality Measures and Aptitude Test Scores

Personality Measures and Aptitude Test Scores	Highs ($N = 46$)	Lows ($N = 43$)	Significance of Difference[a]
Sociability Measures			
Personal Preference Test No. 1	31.7	28.1	NS
Personal Preference Test No. 2	5.9	5.4	NS
Mehrabian's Affiliative Tendency	42.4	42.0	NS
Achievement Motivation Measures			
Mehrabian's Achieving Tendency	28.8	31.8	NS
Empathy Measures			
Knowledge of Prospects Test No. 1	27.9	28.8	NS
Knowledge of Prospects Test No. 2	18.0	18.5	NS
Hogan Empathy Scale	34.8	34.4	NS
Rejection Tolerance Measure			
Social Interactions Test	28.7	28.8	NS
Responsibility-Maturity			
I-E Scale	8.4	7.1	NS
Aptitude Test Scores			
GT (General Technical)	108.4	105.5	NS
CL (Clerical)	108.4	110.9	NS

[a]There were no statistically significant differences, even when applying a very lenient criterion of significance ($p < .10$).

The test instruments not listed in Table 4 were those for which no scoring key existed. These were the tests comprised of items that were thought likely to discriminate between Highs and Lows, but were to be scored empirically. As expected, the vast majority of these items did not discriminate between the High-rated and Low-rated recruiters.

A total of 22 items appeared to discriminate significantly ($p<.10$).[1] These items, which are described below in general terms, are not reproduced in this report. The intent is to minimize the chances that future recruiter applicants might somehow obtain the "correct" answers for the selection test in advance, thereby subverting the purpose of the test. The exact questions and a scoring key will, of course, be made available to USAREC officials and other responsible professionals.

Regarding the 22 discriminating items, it should be kept in mind that whenever a large number of significance tests are made, some ostensibly significant differences may actually be chance differences. Before accepting an item as useful in the selection process, it is necessary to do two things: (a) examine the item's content to decide whether the difference makes sense (if it does not, the difference is especially likely to be a chance result), and (b) subject the item to cross-validation to make sure that it consistently discriminates in the same way.

Four of our 22 items are suspect on the grounds of implausibility, although it is nevertheless possible that they have true validity. Cross-validation will provide a final answer. Another item (mean number of handicaps marked as making his job difficult) is almost certainly a genuinely discriminating item; however, it would be of no use in selecting future recruiters since the question itself does not apply to people without recruiting experience.

The remaining items have at least some plausibility. They pertain to such characteristics as work habits, style of handling finances and debts, educational background, and reactions to challenging or stressful situations.

The question, of course, arises as to whether this set of apparently discriminating items might be useful in actual recruiter selection. A very short test, made up of these 22 items, would be ideal—if it worked. But at this time we cannot be sure that it would; cross-validation would be essential. By this we mean that the set of items must be administered to a new set of recruiters (or perhaps recruiter applicants) and a determination made of how well the items discriminate between the various levels of recruiter talent.

Let us consider a simple paper-and-pencil test since such a test is very economical to administer, score, and interpret. This would eliminate the discriminating item that results from the verbal performance test. The item concerned with the number of recruiting handicaps would also be eliminated since, as explained earlier, it is irrelevant for selection purposes. This would leave 20 items, with possible scores ranging from zero to 20.

Since time and funding limitations did not permit the conduct of a cross-validation phase, we will sketch the following two relatively simple and economical plans for filling the gap.

Plan A. Revised test booklets can be produced and arrangements made with USAREC to administer them to all recruiter applicants who pass the present recruiter selection boards. Probably no more than 10 minutes per subject would be required. Completed test booklets can be transmitted to HumRRO for scoring and retention. After about 200 of the men tested have been on the job for six months, SAS (Simple Achievement Scores) can be computed for each. (These are the scores described in Chapter 2, which express each subject's production over six months as a percentage of the average per recruiter in his DRC.) Correlations can be computed between selection test scores and SAS scores. The resulting data can be analyzed to evaluate the expected value to USAREC of using the test routinely to select recruiters.

[1] Since this item analysis was preliminary, and any apparently significant items would have to be cross-validated, a rather lenient ($p<.10$) criterion of statistical significance was used.

This plan is an example of the follow-up method of selection test validation. It has many advantages, including cost economy, but it does require a rather lengthy period of calendar time, roughly one year.

Plan B. Arrangements can be made with USAREC to distribute copies of the revised test booklet to a large sample consisting of perhaps 200 randomly selected recruiters across the country. Tests can be administered by an officer in each DRC, who will be asked to mail the completed booklets back to HumRRO along with: (a) the six-month production score for each subject, and (b) the average production per recruiter in the DRC. HumRRO researchers will then compute selection test scores and SAS scores, and perform the same analyses described for Plan A.

Plan B has the distinct advantage of requiring little time—perhaps two months. A theoretical disadvantage is that selection test scores might correlate with experience. In other words, some of the things measured in the test might change as a result of recruiting experience. If this were the case, the concurrent validation approach in Plan B would give an erroneous view of how the test would work with true recruiter applicants, who have no recruiting experience. The possibility of such an error can, of course, be checked by obtaining information about the amount of each subject's recruiting experience and examining the correlation between that experience and selection test scores.

WHY WERE SO FEW DISCRIMINATING ITEMS FOUND?

Although this question cannot be answered definitively, it may be useful to examine several possible explanatory factors. For example, the question may be raised as to the adequacy of our two criterion groups—whether they, in fact, represented two extremes of talent. It will be recalled that subjects were selected initially through a supervisory rating system based not only on production records but also on how the subjects operated overall. Twelve of the subjects in the High Criterion Group had made less than 100% of their objective over the six-month period while 13 in the Low-rated group had exceeded their objective. To check the adequacy of our criterion groups, all subjects were reclassified solely on the basis of whether they had met their objectives. In these groups, no significant differences in test performance were found. It thus appears that our basic procedure for forming High and Low groups is superior to the method based solely on the percentage of objectives achieved.

Another possible explanatory factor is that recruiters may already be a highly select group. All must meet certain minimum standards, such as GT score, rank, and time in service. Also, they probably possess similar attributes and, possibly, similar attitudes. The mean number of years on active duty for all men in our sample was 14, and men with this amount of service are likely to have similar philosophies. Hence, it is not surprising that few discriminating items were found in such areas as social interaction attitudes, internal-external orientation, personal opinions and preferences, and personality traits. The few items that did discriminate were mostly from the Background Information Form and Personnel Questionnaire—instruments dealing mainly with matters of fact rather than attitude.

In short, our criterion study results indicate that territorial characteristics account for at least 50% of the variance in production scores. This suggests that individual recruiter characteristics may be relatively unimportant, at least within rather broad limits.

REFERENCES
AND
APPENDICES

REFERENCES

Abrahams, N.M., Neumann, I., and Rimland, B. *Preliminary Validation of an Interest Inventory for Selection of Navy Recruiters*, Research Memorandum, SRM 73-3, Naval Personnel and Training Research Laboratory, San Diego, California, 1973.

Bennett, J.T., and Haber, S.E. *Selection, Deployment, and Evaluation of Marine Recruiters*, The George Washington University School of Engineering and Applied Science, Institute of Management Science and Engineering, Project NR 347020, Office of Naval Research, 1973.

Cravens, David W., and Woodruff, Robert B. "An Approach for Determining Criteria of Sales Performance," in *Journal of Applied Psychology*, vol. 57, no. 3, 1973, pp. 242-247.

Grikscheit, G.M. "An Experimental Investigation of Persuasive Communication in Selling," paper presented at the American Marketing Association Fall Conference, Washington, 1973.

Hogan, R. "Development of an Empathy Scale," in *Journal of Consulting and Clinical Psychology*, vol. 33, no. 3, 1969, pp. 307-316.

McClelland, D.C. *The Achieving Society*, Van Nostrand, Princeton, N.J., 1961.

McMurry, R.N. "The Mystique of Super-Salesmanship," in *Harvard Business Review*, vol. 39, no. 2, March-April, 1961.

Mehrabian, A. "Male and Female Scales of the Tendency to Achieve," in *Educational and Psychological Measurement*, vol. 28, 1968, pp. 493-502.

Mehrabian, A. "Measures of Achieving Tendency," in *Educational and Psychological Measurement*, vol. 29, 1969, pp. 445-451.

Mehrabian, A. "The Development and Validation of Measures of Affiliation Tendency and Sensitivity to Rejection," in *Educational and Psychological Measurement*, vol. 30, 1970, pp. 417-428.

Preston, Joan M., and Gardner, R.C. "Dimensions of Oral and Written Language Fluency," in *Journal of Verbal Learning and Verbal Behavior*, vol. 6, December 1967, pp. 936-945.

Rotter, J.B. "Generalized Expectancies for Internal Versus External Control of Reinforcement," in *Psychological Monographs*, no. 609, vol. 80, no. 1, 1966.

Appendix A

USAREC LETTER DIRECTING RRC'S TO PROVIDE DATA FOR CRITERION STUDY

DEPARTMENT OF THE ARMY
HEADQUARTERS UNITED STATES ARMY RECRUITING COMMAND
FORT SHERIDAN, ILLINOIS 60037

USARCDPA-CS

SUBJECT: Survey Information for HumRRO

Commander, US Army First Recruiting District
Commander, US Army Third Recruiting District
Commander, US Army Fourth Recruiting District
Commander, US Army Fifth Recruiting District
Commander, US Army Sixth Recruiting District

1. The US Army Research Institute, Human Resources Research Organization (HumRRO), is currently conducting a research project for this Command.

2. To supply HumRRO with information they require to complete their research project, the data requested on attached Inclosure 1 must be furnished.

3. Each district will select at random 100 field recruiters and complete the information requested on attached inclosure. To select the 100 recruiters, the last digit of the social security number will be utilized. First, select recruiters whose last digit ends in one. If 100 cannot be selected, then select recruiters whose last digit ends in two. Continue this method until 100 recruiters have been selected.

4. Individuals selected meet the following criteria:

   a. On production in present assignment for the period of July to December 1973.

   b. Have had individual objectives assigned for the same period.

5. The consolidated package containing the 100 survey information sheets will be returned to this headquarters, ATTN: USARCDPA-CS, NLT 28 Feb 74.

FOR THE COMMANDER:

1 Incl
as

W.H. SACHS, JR.
COLONEL, GS
Director of Personnel & Administration

SURVEY INFORMATION SHEET

1. NAME: _____
(LAST) (FIRST) (MIDDLE)

2. SSAN _____ RCID: _____

3. Name & Address of Area Commander:

4. Name & Address of RMS:

5. Name of officer in individual's RMS who may be contacted by HumRRO if assistance is required:

6.	JUL	AUG	SEP	OCT	NOV	DEC
NPS-PS OBJ						
NPS-PS ACC						
% ACC						

7.	JUL	AUG	SEP	OCT	NOV	DEC
POINT SCORE:						

8. Total number of months on production _____.

USAREC Form 344(OT), 19 February 1974.

Appendix B

LIST OF DISTRICT RECRUITING COMMANDS PROVIDING DATA FOR CRITERION STUDY

Northeastern Regional Recruiting Command	
Albany	10
Baltimore	0
Boston	7
Concord	6
Harrisburg	6
Newark	4
Newburg	8
New Haven	4
Ft. Wadsworth	11
Niagara Falls	11
Philadelphia	12
Providence	6
Syracuse	5
Washington	2
Pittsburgh	5
	97

Midwestern Regional Recruiting Command	
Detroit	15
Des Moines	0
Indianapolis	0
Milwaukee	23
Cincinnati	0
Columbus	0
Cleveland	0
Sioux Falls	0
Twin Cities	0
Fargo	0
Omaha	0
St. Louis	0
Lansing	4
Chicago	7
Peoria	16
	66

Southeastern Regional Recruiting Command	
Atlanta	13
Beckley	8
Charlotte	5
Columbia	6
Jacksonville	10
Louisville	13
Miami	6
Montgomery	6
Nashville	8
Raleigh	6
Richmond	11
San Juan	0
	92

Western Regional Recruiting Command	
Boise	4
Helena	2
Honolulu	0
Los Angeles	0
Phoenix	7
Portland	7
Sacramento	0
Fort Douglas	3
Alameda	17
Santa Ana	10
Bellevue	13
	63

Southwestern Regional Recruiting Command	
Albuquerque	7
Amarillo	5
Dallas	9
Denver	0
Houston	9
Jackson	0
Kansas City	15
Little Rock	10
New Orleans	8
Oklahoma City	10
San Antonio	9
	82

Appendix C

FORM USED IN SELECTING SUBJECTS FOR RECRUITER SELECTION RESEARCH

Your Name	Job Title	RMS

DATA COLLECTION FORM
(For Recruiter Selection Research)

The Human Resources Research Organization (HumRRO) is conducting a research project for USAREC. The objective of the research is to develop a procedure for identifying men who are most likely to be both successful and happy in recruiting work. To assist in this project you are being asked to name the 10 recruiters whom you consider to be the very best in your RMS, and also name the 10 whom you believe to be the very poorest in your RMS.

Note Carefully:

(1) Consider only men who have been on production for the past six months.
(2) Do not base your judgement solely on production records or percent of objective achieved. You should also take into account, as best you know how, the quality of the man's territory, any lucky breaks or bad breaks he may have had, and anything else that you think is important. In other words, try to be as fair as you possibly can.
(3) Make your selections all by yourself; do not discuss them with anyone until after you have submitted your completed lists to your RMS Commander.
(4) If you are unable to make 10 selections that you are confident of, just name as many as you can and stop.

TEN BEST (Starting with the very best)	TEN POOREST (Starting with the very poorest)
1. _____ (Last Name, Rank, First Name, Initial)	1. _____ (Last Name, Rank, First Name, Initial)
2. _____	2. _____
3. _____	3. _____
4. _____	4. _____
5. _____	5. _____
6. _____	6. _____
7. _____	7. _____
8. _____	8. _____
9. _____	9. _____
10. _____	10. _____